AD_____

Award Number: DAMD17-99-1-9295

TITLE: Analysis of a Novel 17q25 Cell Cycle Gene Homolog: Is it a Breast Tumor Suppressor Gene?

PRINCIPAL INVESTIGATOR: Linda Kalikin, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: July 2000

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2000	Annual Summary (1 Jul 99 - 30 Jun 00)

4. TITLE AND SUBTITLE
Analysis of a Novel 17q25 Cell Cycle Gene Homolog: Is it a Breast Tumor Suppressor Gene?

5. FUNDING NUMBERS
DAMD17-99-1-9295

6. AUTHOR(S)
Linda Kalikin, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan
Ann Arbor, Michigan 48109-1274

E-MAIL:
lkalikin@umich.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

While a subset of the growth regulating genes involved in breast tumorigenesis has been identified, a complete understanding of all culprit genes is essential for the effective translation of these molecular reagents into successful tools for the medical management of breast cancer. We hypothesize that a 350 kb region on 17q25 detected by our allelic imbalance studies harbors a novel breast tumor suppressor gene. The purpose of this investigation is to characterize the role of this novel gene in normal and tumor development. In the past year, 10 cDNA fragments showing unique expression patterns on multitissue Northern blots have been identified. Extensive characterization of one septin GTPase revealed altered bands on both Southern and Northern blots of breast tumor cell lines. This gene was simultaneously identified as a fusion partner with *MLL* in a leukemia patient and named *MLL* septin-like fusion (*MSF*). Although no functionally significant nucleotide alterations have been found in the MSF coding region to date, *MSF* remains a strong candidate for the breast tumor suppressor gene, and further expression and functional analyses are underway. As well, PCR-based and hybridization-based methods are being used to isolate complete transcripts from other cDNA fragments in the region.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

LMV Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

LMV Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

LMV For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

7/28/00

PI - Signature Date

Table of Contents

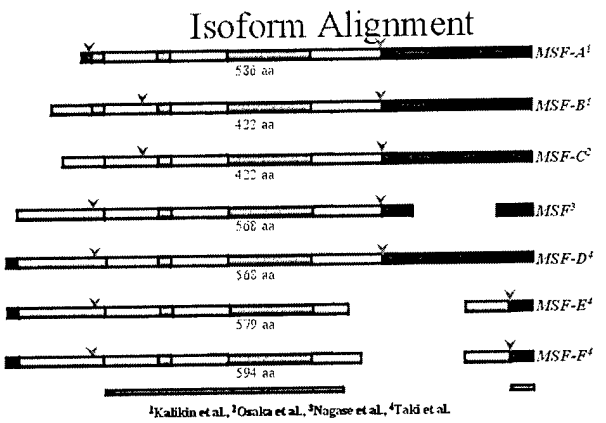
INTRODUCTION

Human chromosome 17q25.1-17q25.2 is altered in a variety of solid and hematological malignancies, suggesting the location of at least one novel cancer gene. We defined an interval of allelic imbalance in this region by PCR using short tandem repeat polymorphisms on over 70 matched breast normal/tumor paraffin-embedded microdissected tissue samples (Kalikin *et al.*, 1996, 1997). Additional studies provided supportive evidence for the presence of a novel breast tumor suppressor gene in this area (Theile *et al.*, 1995; Phelan *et al.*, 1998). Construction of a BAC/PAC/P1 genomic contig through the approximately 350 kb region allowed more precise mapping of genes and ESTs and the isolation of expressed sequences (Kalikin *et al.*, 1999). The purpose of this project is to identify this novel suppressor gene and to define its role in normal and breast tumor development. To accomplish this, we proposed to analyze candidate genes in breast tumor cell lines and matched normal and breast tumor pairs by RT-PCR and Northern hybridization. Mutational analysis would be carried out in samples with aberrant expression levels or transcript sizes. Further expression characterization would include multitissue Northern blot analyses and *in situ* hybridization. Functional characterization of candidate genes for which mutations have been identified in breast tumors would be by transfection into breast cancer cell lines and immunocompromised mice. Yeast two-hybrid would be used to identify potential protein-protein interactions.

BODY

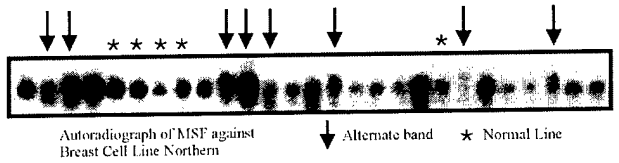
A. Research

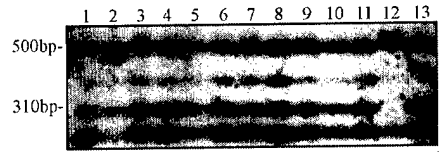
As proposed in our approved Statement of Work, the first year of this grant has focused on further characterization of a novel septin GTPase cDNA fragment as the candidate tumor suppressor gene. Septins belong to a highly conserved gene family that localize to the cleavage furrow in yeast (review Field *et al.*, 1999) and to the contractile ring in animal cells (review Sanders and Fields, 1994). Septins have been shown to cause cell-cycle arrest and impaired cytokinesis when mutated in yeast and to result in multinucleated cells when mutated in animal cells. They polymerize into filaments (Field and Kellogg, 1996; Frazier *et al.*, 1998), exhibit GTPase activity (Field and Kellogg, 1996), coordinate cell cycle progression by binding to the mitosis-inducing protein kinases HSL1, KCC4, and GIN1 (yeast; Barral *et al.*, 1999) and bind in a GDP-associated form to membrane phospholipids (animal cells; Zhang *et al.*, 1999). The initial 339 bp septin fragment was isolated by solution hybrid capture using cDNA derived from mammary gland poly(A)$^+$ RNA and contained an open reading frame spanning its entirety. Direct hybridization to a λGT11 breast cDNA library, RACE, and NCBI database *in silico* walking were utilized to isolate the full open reading frame. PCR amplification in normal epithelial cell line cDNA and subsequent sequencing confirmed correct gene sequence assembly. These efforts yielded two transcripts of 3737 bp and 3970 bp that differed only at their initial 36 nucleotides and 269 nucleotides respectively (Kalikin *et al.*, 2000). Concurrent with this work, this gene was identified as the carboxy-terminal partner in a t(11;17)(q23;q25) fusion protein with *MLL* (mixed lineage leukemia) in acute myeloid leukemia (AML) and was named *MSF* (MLL septin-like fusion; Osaka *et al.*, 1999). *MSF* differed from our two variants over the first 776 bp and was lacking 1642 nucleotides at the 3' end. We designated the smaller of our *MSF* variants *MSF-A* (Accession No. AF189713) and the larger one *MSF-B* (Accession No. AF189712). A fourth transcript (KIAA0991; Accession No. AB023208) independently mapped to chromosome 17 (Nagase *et al.*, 1999) was unique over the first 317 bp compared to the other three variants and included the 1642 nt at the 3' end in our two variants. Within our lab, we called KIAA0991 (3938 bp) *MSF-C*. A later report describing a t(11;17)(q23;q25) AML patient identified three other *MSF* variant transcripts isolated from a normal immortalized cell line (Taki *et al.*, 1999). All three transcripts were identical to *MSF* at the 5' end but with the addition of 37 bp before the first *MSF* nucleotide. One transcript matched our 3' sequence while the other two had different 3' deletions.



The 5' and 3' sequence differences in the *MSF* variants resulted in distinct open reading frames, all of which began at methionines within a consensus Kozak sequence and terminated at a stop codon upstream of an AATAAA polyadenylation site. The MSF-A protein was predicted to be of 586 amino acids. The first 20 residues of MSF-A differed from the first 7 residues of the 568 amino acid MSF protein after which the proteins were identical (see Fig. Isoform Alignment). MSF-B and MSF-C encoded identical predicted 422 residue proteins whose sequences were found in-frame of both larger protein products from MSF and MSF-A. Of the protein products predicted by Taki *et al.*, two matched MSF except for the last 34 residues (MSF-E; 594 amino acids) and last 18 residues (MSF-F; 579 amino acids) which were replaced by 44 unique amino acids. The third protein product was identical to MSF. All protein products contained a highly conserved GTPase domain. In addition to binding and hydrolyzing GTP, proteins with this domain have been shown to transmit membrane signals, direct protein synthesis and control cellular proliferation and differentiation (review Bourne *et al.*, 1991). A xylose isomerase 1 domain was also identified in all isoforms. This sequence ([LI]-E-P-K-P-x(2)-P) has previously been recognized only in microorganisms using the sugar interconverting enzyme xylose isomerase and is thought to be necessary for catalysis and cation ligand binding (Dauter *et al.*, 1989; Henrick *et al.*, 1989). It is unclear what the function of such a domain would be in advanced organisms. A domain search of the sequence databases identified the FYN-binding protein (FYB), also known as the SLP-76 associated protein (SLP-130), as the only other human protein to contain this domain. FYB is expressed specifically in T cells and myeloid cells and has been shown to be involved in FYN and SLP-76 signaling cascades (da Silva *et al.*, 1997; Musci *et al.*, 1997).

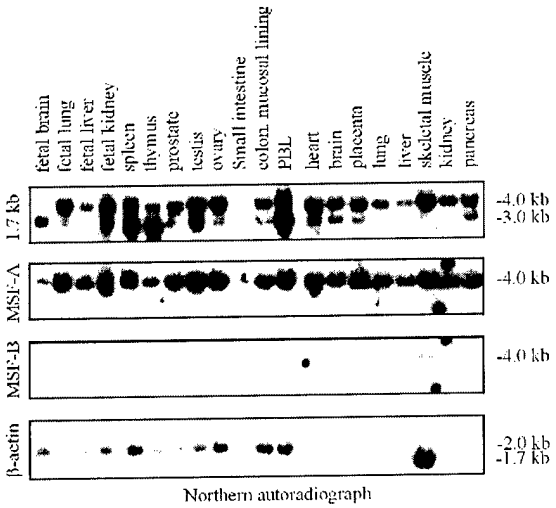
Thus, with its high degree of conservation to the septin family and with the presence of the highly conserved GTPase domain known to be involved in cellular proliferation, *MSF* is an extremely exciting candidate for the breast suppressor gene. As proposed in this grant, we would initially analyze such candidate genes by looking for altered expression levels or aberrant transcript sizes on Northern blots of breast normal and tumor samples. A 1.7 kb fragment common to all *MSF* transcripts was hybridized to a breast normal and tumor cell line total RNA Northern blot. Eight of 22 tumor cell lines exhibited apparent alternative transcript bands not found in 5 normal breast cell lines (see Fig. Autoradiograph of MSF against Breast Cell Line Northern). Hybridization to a Southern blot of *MspI* digested breast normal and tumor cell line DNA revealed rearranged and missing bands in two additional tumor cell lines (see Fig. MspI Southern Blot, lanes 2, 12). Sequencing through RT-PCR products amplified from these tumor cell lines is approximately 90% complete and has only identified silent polymorphisms with analysis of the unique 5' sequences from each variant currently incomplete. Similarly no functionally significant nucleotide alterations have been found in the MSF coding region of 7 paraffin embedded breast tumors with minimal 17q25 allelic imbalance.



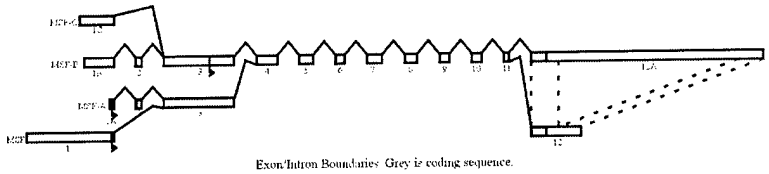


MspI Southern Blot

These initial results do not support a direct role for *MSF* mutations in breast tumorigenesis. However, a role for alternatively spliced transcript expression has not been excluded. Therefore, further MSF expression and functional analyses are underway. RNA expression studies using the *MSF* 1.7 kb probe showed differential expression of 3.0 kb and 4.0 kb transcripts in all adult and fetal tissues tested (see Fig. Northern autoradiograph). These transcripts appear to be developmentally regulated given that their ratio of expression levels changes from fetal brain to adult brain and from fetal kidney to adult kidney. A probe spanning sequence unique to *MSF-A* detected specific expression of the 4.0 kb transcript in all tissues. Another probe unique to *MSF-B* detected a 4.0 kb transcript in only skeletal muscle, although results from a control probe suggest that this lane may be overloaded. Sequencing from BACs spanning the candidate region identified 9 common exons and 3 alternatively spliced exons ranging in size from 39-291 bp (see Fig. Exon/Intron Boundaries). It is unclear as to the origin of the two predicted proteins that vary in the 3' region in the Taki *et al.* paper as we have not detected any splice sites in this interval. Based on this genomic structure, we have determined that to date all the leukemic rearrangements between *MLL* and *MSF* occur in-frame at the splice junction between exon 1 and exon 3. Interestingly, the identification of *MSF* as a fusion protein partner with *MLL* in multiple leukemia patients suggests that *MSF* may have activating potential. Its localization to a discrete region of loss in breast tumors, and the suggestion of altered *MSF* bands on breast tumor cell line Northern and Southern blots supports its role as a candidate tumor suppressor gene. These potentially opposite mechanisms for a single disease gene are not unprecedented and include the *RET* gene that is subject to both activating and inactivating mutations which lead to distinct clinical manifestations (Eng, 1999). Functional characterization of MSF protein-protein interactions is underway by yeast two-hybrid. As an initial step, proteins of predicted weights have been expressed from initial TA vector constructs harboring MSF-A (65.4 kDa) and MSF-B (47.5 kDa) coding regions (see Fig. MSF Expression).



Northern autoradiograph


Exon/Intron Boundaries Grey is coding sequence.

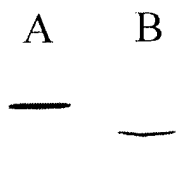
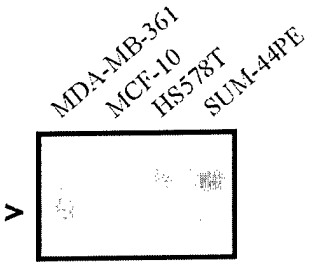


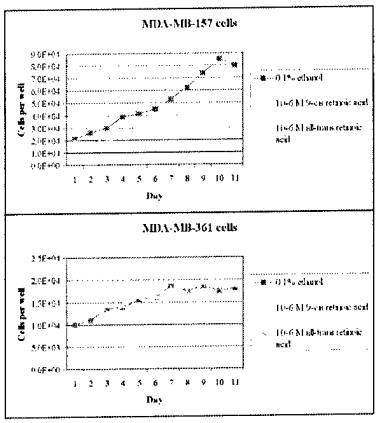
Fig. MSF Expression

Although *MSF* may be interesting in leukemias or may be a novel component of the cell cycle, it remains possible that our completed results will not be supportive of *MSF* as the 17q25 suppressor gene, especially if no mutations are identified in breast tumors. Thus, we have continued to examine other cDNAs from this region. In our preliminary results for this grant, we had screened genes and ESTs generally localized to distal 17q24-proximal 17q25 by STS-PCR against contig P1 and BACs to identify potential positional candidate genes. Of 29 genes and 10 ESTs tested, only *SEC14L1*, which encodes a protein with partial homology to a yeast secretory protein and a squid retinal-binding protein, mapped to the critical interval. As retinoic acid has been shown to inhibit the proliferation of certain breast cancers (Wilcken *et al.*, 1996; Van heusden *et al.*, 1998), loss of function of a gene with retinal binding properties could serve to augment the progression of mammary tumorigenesis. Hybridization of a partial *SEC14L1* cDNA fragment to a breast tumor and

normal cell line Northern blot revealed a potentially altered band in at least one tumor cell line (MDA-MB-361) out of 21 (see Fig. SEC14L1 Northern Autoradiograph).



SEC14L1 Northern Autoradiograph



Retinoic Acid Effects on Cell Growth

Preliminary experiments to investigate the proliferative affects of retinoic acid isomers on this cell line suggested that its growth rate was unaffected by the presence of retinoic acid while a control breast cancer cell line MDA-MB-157 showed reduced growth rates as expected (See Fig. Retinoic Acid Effects on Cell Growth). However no functionally significant sequence variations were identified in the coding region. Experiments are on-going to sequence the 5' and 3' regulatory regions and to investigate protein expression.

Concurrent with *MSF* and *SEC14L1* analyses, we continue to identify other cDNAs in the candidate region. Over the past year, we mapped by STS-PCR 10 ESTs derived from the databases. Solution hybrid capture was utilized to isolate novel expressed sequences. Of approximately 9 kb of sequence generated from 40 cDNA fragments, only hits for *SEC14L1* and *MSF* (excluding ribosomal and repeat sequences) have been identified from NCBI Genbank BLAST analyses, despite this past year's vast sequencing efforts to decode the human genome. Six database-derived ESTs and 4 solution hybrid capture-derived cDNAs showed strong signals on multitissue Northerns (see Table cDNA Tissue Expression; solution hybrid capture clones are in gray). Several cDNA fragments did not reveal transcripts on any of the multitissue Northern blots. One possibility is that these represent genes that are expressed in a very narrow tissue-specific or developmental-specific range. These fragments will be hybridized against improved multitissue Northern blots with an increased selection of tissues that are now available from Clontech. Note that while solution hybrid capture clones were isolated from mammary gland poly(A)+ RNA and other ESTs may also be expressed in normal mammary gland, this tissue was not available on the original multitissue Northern blots and so are not listed in the cDNA Expression Table. Mammary gland RNA will be a lane on these new blots.

cDNA TISSUE EXPRESSION

CLONE	TISSUE	SIZE (KB)	CLONE	TISSUE	SIZE (KB)
1A	fetal liver	2.0, 3.0	1C3A2	spleen	7.0
	skeletal muscle	4.4, 7.5		thymus	7.0
	liver	2.4		testis	4.4
	placenta	1.2		peripheral blood leukocytes	7.0
AI692509	heart	1.3, 2.0	H49244	fetal liver	1.3
	pancreas	1.3, 2.0, 2.5		bone marrow	1.3
	placenta	1.3		thyroid	9.5
3L	liver	2.0	1A10R4	ubiquitous	2.0
AA287995	fetal brain	6.0	AI906327	fetal liver	6.5
	brain	6.0		placenta	6.5
	testis	6.0		skeletal muscle	1.0, 7.0
T95297	fetal brain	5.5	T78018	fetal brain	3.0
	fetal kidney	5.5		brain	3.0
	brain	5.5		others	variable?
	spleen	5.5			

We have employed multiple methods to isolate complete transcripts including the following: direct hybridization to an ovarian, a testis and two different mammary gland phage cDNA libraries immobilized on nylon membranes; direct hybridization to Human Universal cDNA Library (HUCL) arrays (Stratagene) with over 290,000 clones constructed from 29 tissues and averaging 1.7 kb inserts; PCR-based methods of RACE (rapid amplification of cDNA ends; Clontech), SMART (switching mechanism at 5' end of RNA template; Clontech), and SPICE (system for PCR identification of cDNA ends; (Starrs and Davies, 1999); and *in*

silico subcloning through Genbank (http://www.ncbi.nlm.nih.gov/BLAST/); DoubleTwist (https://www.doubletwist.com) and Stratagene http://www.stratagene.com/gc/clone.htm). No method has been appreciably more successful than any other method. Inexplicably, all methods failed to produce additional gene sequences on a number of cDNA fragments, despite observing distinct RNA expression bands on multitissue Northerns. While we have found that smaller cDNA fragments give clear signals on the poly(A)+ multitissue Northern blots such as those 300-500 bp fragments isolated by solution hybrid capture, larger fragments are necessary for interpretable data against our in-house total RNA breast tumor cell line Northern blots. Therefore, *in silico* sequence searches are continuing and additional methods are being employed before direct candidate analysis on these breast tumor cell line blots can be completed.

Future studies include:

- completion of *MSF* sequencing in breast tumor cell lines with apparent rearranged Southern and Northern bands.
- subcloning and sequencing of *MSF MspI* rearranged breast tumor cell line fragments.
- generation of MSF polyclonal antibodies to investigate protein expression levels in breast tumor cell lines.
- transfection of SEC14L1 protein expression construct into MDA-MB-361 and repeat of retinoic acid experiments.
- continued cloning and candidate testing of full length novel cDNAs within the candidate region.
- use of available genomic sequence through the candidate interval, now estimated to be 80% from in-house BAC sequencing and database entries, to aid in gene identification using especially exon predictor programs.

B. Training

My training has adhered quite closely to that as proposed in the grant. The University of Michigan Medical Center ranks among the top facilities in the country and provides a stimulating and exciting environment for cancer genetic research and postdoctoral training. I attended weekly laboratory meetings with my mentor Dr. Elizabeth Petty's and my co-mentor Dr. Eric Fearon's research groups. These meetings allowed for critical comments on my data and future plans. I presented these data at the 1999 Department of Human Genetics Annual Retreat, the 2000 Internal Medicine Research Day and the 1999 American Association for Cancer Research Annual Meeting. Based on my research from this past year, I have been awarded a University of Michigan Comprehensive Cancer Institutional grant from the American Cancer Society to begin August 2000 to investigate further the role of MSF by comparative genomics in yeast. In addition, I regularly attended the weekly Cancer Center grand rounds seminar, the monthly Cancer Center journal club, and the monthly mouse journal club as well as other pertinent seminars as available. This coming year I will be presenting my work at the 2000 American Society of Human Genetics Annual Meeting.

REFERENCES

Barral Y, Parra M, Bidlingmaier S, Snyder M. Nim1-related kinases coordinate cell cycle progression with the organization of the peripheral cytoskeleton in yeast. Genes Dev 1999;13:176-87.

Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: conserved structure and molecular mechanism. Nature 1991;349:117-27.

da Silva AJ, Li Z, de Vera C, Canto E, Findell P, Rudd CE. Cloning of a novel T-cell protein FYB that binds FYN and SH2-domain-containing leukocyte protein 76 and modulates interleukin 2 production. Proc Natl Acad Sci U S A 1997;94:7493-8.

Dauter Z, Dauter M, Hemker J, Witzel H, Wilson KS. Crystallisation and preliminary analysis of glucose isomerase from Streptomyces albus. FEBS Lett 1989;247:1-8.

Eng C. RET proto-oncogene in the development of human cancer. J Clin Oncol 1999;17:380-93.

Field CM, al-Awar O, Rosenblatt J, Wong ML, Alberts B, Mitchison TJ. A purified Drosophila septin complex forms filaments and exhibits GTPase activity. J Cell Biol 1996;133:605-16.

Field CM, Kellogg D. Septins: cytoskeletal polymers or signalling GTPases? Trends Cell Biol 1999;9:387-94.

Frazier JA, Wong ML, Longtine MS, Pringle JR, Mann M, Mitchison TJ, Field C. Polymerization of purified yeast septins: evidence that organized filament arrays may not be required for septin function. J Cell Biol 1998;143:737-49.

Henrick K, Collyer CA, Blow DM. Structures of D-xylose isomerase from Arthrobacter strain B3728 containing the inhibitors xylitol and D-sorbitol at 2.5 A and 2.3 A resolution, respectively. J Mol Biol 1989;208:129-57.

Kalikin LM, Qu X, Frank TS, Caduff RF, Svoboda SM, Law DJ, Petty EM. Detailed deletion analysis of sporadic breast tumors defines an interstitial region of allelic loss on 17q25. Genes Chromosomes Cancer 1996;17:64-8.

Kalikin LM, Frank TS, Svoboda-Newman SM, Wetzel JC, Cooney KA, Petty EM. A region of interstitial 17q25 allelic loss in ovarian tumors coincides with a defined region of loss in breast tumors. Oncogene 1997;14:1991-4.

Kalikin LM, George RA, Keller MP, Bort S, Bowler NS, Law DJ, Chance PF, Petty EM. An integrated physical and gene map of human distal chromosome 17q24- proximal 17q25 encompassing multiple disease loci. Genomics 1999;57:36-42.

Kalikin LM, Sims HL, Petty EM. Genomic and expression analysis of alternatively spliced transcripts of the *MLL* septin-like fusion gene *(MSF)* that map to a 17q25 region of loss in breast and ovarian tumors. Genomics 2000;63:165-72.

Musci MA, Hendricks-Taylor LR, Motto DG, Paskind M, Kamens J, Turck CW, Koretzky GA. Molecular cloning of SLAP-130, an SLP-76-associated substrate of the T cell antigen receptor-stimulated protein tyrosine kinases. J Biol Chem 1997;272(18):11674-7.

Nagase T, Ishikawa K, Suyama M, Kikuno R, Hirosawa M, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 1999;6:63-70.

Osaka M, Rowley JD, Zeleznik-Le NJ. *MSF* (MLL septin-like fusion), a fusion partner gene of *MLL*, in a therapy-related acute myeloid leukemia with a t(11;17)(q23;q25). Proc Natl Acad Sci USA 1999;96:6428-33.

Phelan CM, Borg A, Cuny M, Crichton DN, Baldersson T, Andersen TI, Caligo MA, Lidereau R, Lindblom A, Seitz S, et al. Consortium study on 1280 breast carcinomas: Allelic loss on chromosome 17 targets subregions associated with family history and clinical parameters. Cancer Res 1998;58:1004-12.

Sanders SL, Field CM. Cell division. Septins in common? Curr Biol 1994;4:907-10.

Starrs S, Davies RJH. System for PCR identification of cDNA ends (SPICE). BioTech 1999;27:46-8.

Taki T, Ohnishi H, Shinohara K, Sako M, Bessho F, Yanagisawa M, Hayashi Y. *AF17q25*, a putative septin family gene, fuses the *MLL* gene in acute myeloid leukemia with t(11;17)(q23;q25). Cancer Res 1999;59:4261-5.

Theile M, Hartman S, Scherthan H, Arnold W, Deppert W, Frege R, Glaab RF, Haensch W, Scherneck S. Suppression of tumorigenicity of breast cancer cells by transfer of chromosome 17 does not require transferred BRCA1 and p53 genes. Oncogene 1995;10:439-47.

Van heusdan J, Wouters W, Ramaekers FCS, Krekels MDWG, Dillen L, Borgers M, Smets G. All-*trans*-retinoic acid metabolites significantly inhibit the proliferation of MCF-7 human breast cancer cells in vitro. Br J Cancer 1998;77:26-32.

Wikcken NRC, Sarcevic B, Musgrove EA, Sutherland RL. Differential effects of retinoids and antiestrogens on cell cycle progression and cell cycle regulatory genes in human breast cancer cells. Cell Growth Diff 1996;7:65-74.

Zhang J, Kong C, Xie H, McPherson PS, Grinstein S, Trimble WS. Phosphatidylinositol polyphosphate binding to the mammalian septin H5 is modulated by GTP. Curr Biol 1999;9:1458-67.

APPENDIX

1) Key research accomplishments:
   - isolation of *MSF* (*MLL* septin-like fusion) from the 17q25 candidate breast tumor suppressor gene region; genomic and expression analyses of *MSF* alternatively spliced variants.
   - identification of rearranged *MSF* bands on Southern and Northern blots of breast tumor cell lines; no functional sequence alterations found to date.
   - identification of breast tumor cell line with an apparent altered *SEC14L1* transcript; this cell line growth rate did not decrease after treatment with retinoic acid derivatives, unlike a control breast cancer cell line.
   - identification of 10 cDNAs to the candidate region with unique multitissue expression patterns.
   - generation and assembly of 80% of genomic sequence through the 350 kb candidate region.

2) Reportable outcomes:
   - Kalikin LM, Sims HL, Petty EM. Genomic and expression analyses of alternatively spliced transcripts of the *MLL* septin-like fusion gene (*MSF*) that map to a 17q25 region of loss in breast and ovarian tumors. Genomics 1999;63:165-72.
   - awarding of a University of Michigan Comprehensive Cancer Center Institutional Grant from the American Cancer Society entitled "Functional characterization in yeast of human septin MSF, a 17q25 cancer-associated gene" to begin August 1, 2000.

3) Please find included 3 reprints of Kalikin *et al.*, 2000.

Genomics 63, 165–172 (2000)
doi:10.1006/geno.1999.6077, available online at http://www.idealibrary.com on IDEAL®

Genomic and Expression Analyses of Alternatively Spliced Transcripts of the *MLL* Septin-like Fusion Gene (*MSF*) That Map to a 17q25 Region of Loss in Breast and Ovarian Tumors

Linda M. Kalikin, Heather L. Sims, and Elizabeth M. Petty[1]

Department of Internal Medicine and Department of Human Genetics, The University of Michigan, Ann Arbor, Michigan 48109-0638

Received September 1, 1999; accepted November 23, 1999

We previously defined a common region of 17q25 loss in breast and ovarian tumors, suggesting localization of at least one putative tumor suppressor gene. Genomic clones from the interval were used to isolate candidate transcripts. One novel transcript had strong homology to a septin family of GTPase genes involved in cytokinesis. This gene was recently identified as a myeloid/lymphoid leukemia (*MLL*) fusion protein partner in acute myeloid leukemia and was named *MSF* (*MLL* septin-like fusion). As this gene may play roles in both leukemogenesis and tumorigenesis, it is essential to understand its structure and normal expression. We cloned two human alternative transcripts and identified a third database variant of *MSF*. RNA expression studies with a probe common to the three novel sequences showed differential expression of 4.0- and 3.0-kb transcripts in all adult and fetal tissues tested. A probe spanning sequence unique to one *MSF* variant detected specific expression of the 4.0-kb transcript in all tissues. Another probe unique to a different *MSF* variant detected a 4.0-kb transcript only in skeletal muscle. Proteins of 422 and 586 amino acids were predicted from the novel alternate transcripts and included both a xylose isomerase 1 domain and a GTPase domain. Nine common exons, three alternatively spliced exons, and six polymorphisms were identified. © 2000 Academic Press

INTRODUCTION

Tumorigenesis is widely accepted to be a multistep process. Based on studies in colon cancer, the progression from early adenoma to carcinoma occurs as mutations accumulate in an increasing number of genes and leads to deregulated cellular growth and proliferation (Fearon, 1997). Characterization of these pathways will be important to understanding the pathogenesis and facilitating the management of cancer. To identify additional genes involved in breast and ovarian carcinogenesis, we conducted loss of heterozygosity (LOH) studies in breast and ovarian tumors and defined the location of a putative 17q25 tumor suppressor gene (Kalikin *et al.*, 1996, 1997). Additional studies (Saito *et al.*, 1993; Theile *et al.*, 1995; Phelan *et al.*, 1998) provided supportive evidence for the presence of this suppressor locus.

Construction of a genomic contig though the interval allowed more precise mapping of genes and ESTs (Kalikin *et al.*, 1999), and contig clones were also utilized to isolate expressed sequences. One novel transcript revealed a high level of sequence homology to the septin subfamily of GTPase genes. GTPases have a highly conserved domain that binds and hydrolyzes GTP (Bourne *et al.*, 1991). Members of this protein family include those that transmit membrane signals, direct protein synthesis, and control cellular proliferation and differentiation. The septin subfamily includes *CDC10, NEDD5,* and *H5* (Kinoshita *et al.*, 1997). Localization of these genes to the contractile ring in mammalian cells and to the cleavage furrow in yeast suggests that they are important in regulating cytokinesis. Recently, this 17q25 septin-like gene was identified as part of a fusion protein with the myeloid/lymphoid leukemia gene (*MLL*) in a therapy-induced acute myeloid leukemia (t-AML) patient with a t(11;17)(q23; q25) rearrangement and was named *MSF* (*MLL* septin-like fusion; Osaka *et al.*, 1999). *MLL* on 11q23 provides the N-terminal portion of an in-frame fusion protein with at least 19 other distinct genes in leukemia patients with 11q23 translocations (Rowley, 1998; Osaka *et al.*, 1999). While *MLL* appears to play a causal role in leukemogenesis, evidence suggests that the C-terminal fusion partner also is important in the hematopoetic transformation (Corral *et al.*, 1996). Thus, the simultaneous identification of *MSF* as an *MLL* fusion protein partner in a leukemia patient (Osaka *et al.*, 1999) and as an attractive positional candidate breast and ovarian tumor suppressor gene

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. (*MSF*-A) AF189713 and (*MSF*-B) AF189712.

[1] To whom correspondence and reprint requests should be addressed. Telephone: (734) 764-1549. Fax: (734) 647-7979. E-mail: epetty@umich.edu.

0888-7543/00 $35.00

based on our LOH analyses makes further characterization of the structure and expression of this gene essential to understanding its potential roles in leukemogenesis and in solid tumorigenesis.

MATERIALS AND METHODS

Isolation and assembly of transcript sequences. cDNA clones prepared from normal mammary poly(A)$^+$ RNA (Clontech Laboratories) that mapped to genomic clones in the 17q25 candidate region were identified by solution hybrid capture as described (Futreal *et al.,* 1994). Digested genomic clone DNA was ligated to Uni-Amp *Sal*I adaptors (5′-CCTCTGAAGGTTCCAGAATCGATAGGTCGACCG-3′ and 5′-PO_4CGGTCGACCTATCGATTCTGGAACCTTCAGAGGTTT-3′, Clontech Laboratories) and amplified by PCR using a biotinylated Uni-Amp primer (5′-biotin CCTCTGAAGGTTCCAGAATCGATAG-3′, Clontech Laboratories). Reactions were purified through a QIAquick column (Qiagen Inc.). Captured cDNAs were TA subcloned into pCRII-TOPO (Invitrogen Corp.). Additional sequence surrounding the septin-like solution hybrid capture cDNA fragment was obtained by direct hybridization to a λgt11 breast cDNA library (Swaroop and Xu, 1993) immobilized on Hybond-N$^+$ membrane (Amersham). 3′ sequences were obtained by EST database walking (http://www.ncbi.nlm.nih.gov/blast). 5′ sequences were isolated by RACE from a Marathon-Ready human mammary gland cDNA library (Clontech Laboratories) using nested gene-specific primers R3 (5′-CACCTGCTTGGACGAGATGTCAATGG-3′) and R4 (5′-GGAGCGTTGGCTTAGGGAGTCCACAT-3′) with adaptor-specific primers AP1 and AP2, respectively, according to the manufacturer's instructions (Clontech Laboratories). Amplified fragments were TA subcloned into pCRII-TOPO (Invitrogen Corp.).

Southern and Northern analyses. BAC and P1 clones in the 17q25 candidate region have been previously described (Kalikin *et al.,* 1999). DNA was purified from 500-ml cultures through Qiagen-tip 500 columns using a modified Plasmid Maxi Kit protocol available from the manufacturer (Qiagen Inc.). Briefly, these modifications included increasing the volumes of Buffers P1, P2, and P3 to 50 ml each and eluting the DNA from the column in five aliquots of 2 ml Buffer QF prewarmed to 65°C. DNA was digested with *Not*I and separated on a 1.0% SeaKem agarose gel (FMC-BioProducts) in 0.5× TBE on a Bio-Rad Chef Mapper at 14°C with parameters 6 V/cm, 2-s initial switch time, 10-s final switch time, linear ramping factor, 120° angle, and 12-h run time. Gels were immobilized on Hybond-N$^+$ (Amersham). Multitissue Northern blots were purchased from Clontech Laboratories. Probes were generated by random priming (Feinberg and Vogelstein, 1983) with [α-^{32}P]dCTP and were hybridized to filters at 60°C in 0.5 M NaH_2PO_4, 7% SDS, 1 mM EDTA buffer (Southern blots; Church and Gilbert, 1984) or ExpressHyb (Northern blots; Clontech Laboratories). Filters were washed twice in 2× SSC/0.1% SDS and once in 0.1× SSC/0.1% SDS and exposed overnight on Kodak X-Omat AR film.

Sequencing. cDNA plasmid clones and PCR products were sequenced by The University of Michigan Sequencing Core. BAC DNA was sequenced using a modification of the manufacturer's Thermo Sequenase cycle sequencing protocol (USB Corp.). Template DNA was increased to 2 μg per reaction and primer to 2.5 pmol. Termination master mix and labeled ddNTPs were all doubled in volume for each reaction.

RT-PCR. Total RNA was isolated by Trizol (Gibco BRL Life Technologies) from cultured human epithelial cells. cDNA was synthesized using dT and random primers with SuperScript II reverse transcriptase (Gibco BRL). PCR was performed in 50 mM KCl, 10 mM Tris–HCl, pH 9.0, 0.1% Triton X-100, 1.5 mM $MgCl_2$ with each dNTP at 200 μM, each primer at 0.5 μM, and 1 unit of *Taq* Polymerase (Promega Corp.) or using the Expand Long Template PCR System with Buffer 1 (Boehringer Mannheim Corp.).

RESULTS AND DISCUSSION

Identification of Alternative MSF Transcripts

A candidate 17q25 suppressor gene interval defined by LOH spanned an estimated 3 cM (Kalikin *et al.,* 1997, 1999). LOH analysis on breast and ovarian tumors with additional polymorphic microsatellite markers further narrowed the minimal candidate region to less than 500 kb surrounding D17S937. In an effort to identify candidate tumor suppressor genes, cDNA fragments were isolated from BACs and P1s mapping to the smallest common interval of loss. One cDNA clone isolated by solution hybrid capture (Futreal *et al.,* 1994) from BAC 334m6 (Kalikin *et al.,* 1999) showed high conservation to a septin-like GTPase gene family (Bourne *et al.,* 1991; Kinoshita *et al.,* 1997) and contained an open reading frame (ORF) spanning the entire fragment. To determine the sequence of the full open reading frame, the 339-bp captured cDNA was used to screen a λGT11 breast cDNA library from which a 2.1-kb cDNA with a complete GTPase domain was isolated. Additional 5′ sequence was isolated by RACE using a normal mammary gland Marathon-Ready cDNA library (Clontech Laboratory), and 3′ sequence was obtained by database EST walking. cDNA sequence was confirmed by amplification and sequencing in normal epithelial cell line cDNA. Sequence comparisons revealed a high degree of internal homology to the recently published *MSF* (GenBank Accession No. AF123052) with unique sequences at the 5′ and 3′ ends. At the 5′ end, our RACE efforts yielded both 276- and 610-bp amplified fragments. These products had 183 bp in common with *MSF* and an additional 59 bp in common with each other before deviating with unique 36- and 269-bp 5′ ends (Fig. 1A). At the 3′ end, our transcripts contained an additional 1642 nucleotides that were absent in the *MSF* sequence. However, our 3′ sequence was identical except for two bases and a shortened polyA tail to that of KIAA0991 (Accession No. AB023208), a transcript independently mapped to chromosome 17 (Nagase *et al.,* 1999) with a high level of homology to *MSF* (Fig. 1A). KIAA0991 5′ sequence was distinct from *MSF* and our derived sequences. Thus, we have identified two novel alternative transcripts of *MSF* that we designated *MSF*-A (3737 bp) and *MSF*-B (3970 bp). These sequences were deposited with the GenBank/EMBL database under Accession Nos. AF189713 and AF189712, respectively. KIAA0991, a third apparent *MSF* variant (3938 bp), is referred to as *MSF*-C in this study. As it is not uncommon for key elements controlling temporal- and tissue-specific expression to be located in the 5′ and 3′ flanking regions of a transcript (Jackson, 1993; Jean *et al.,* 1999), the identification of these multiple *MSF* variants suggests that the expression of *MSF* may involve complex regulation in different tissues.

Prediction of Protein Structure and Function

The 5′ sequence variability in the *MSF* alternate transcripts resulted in distinct open reading frames

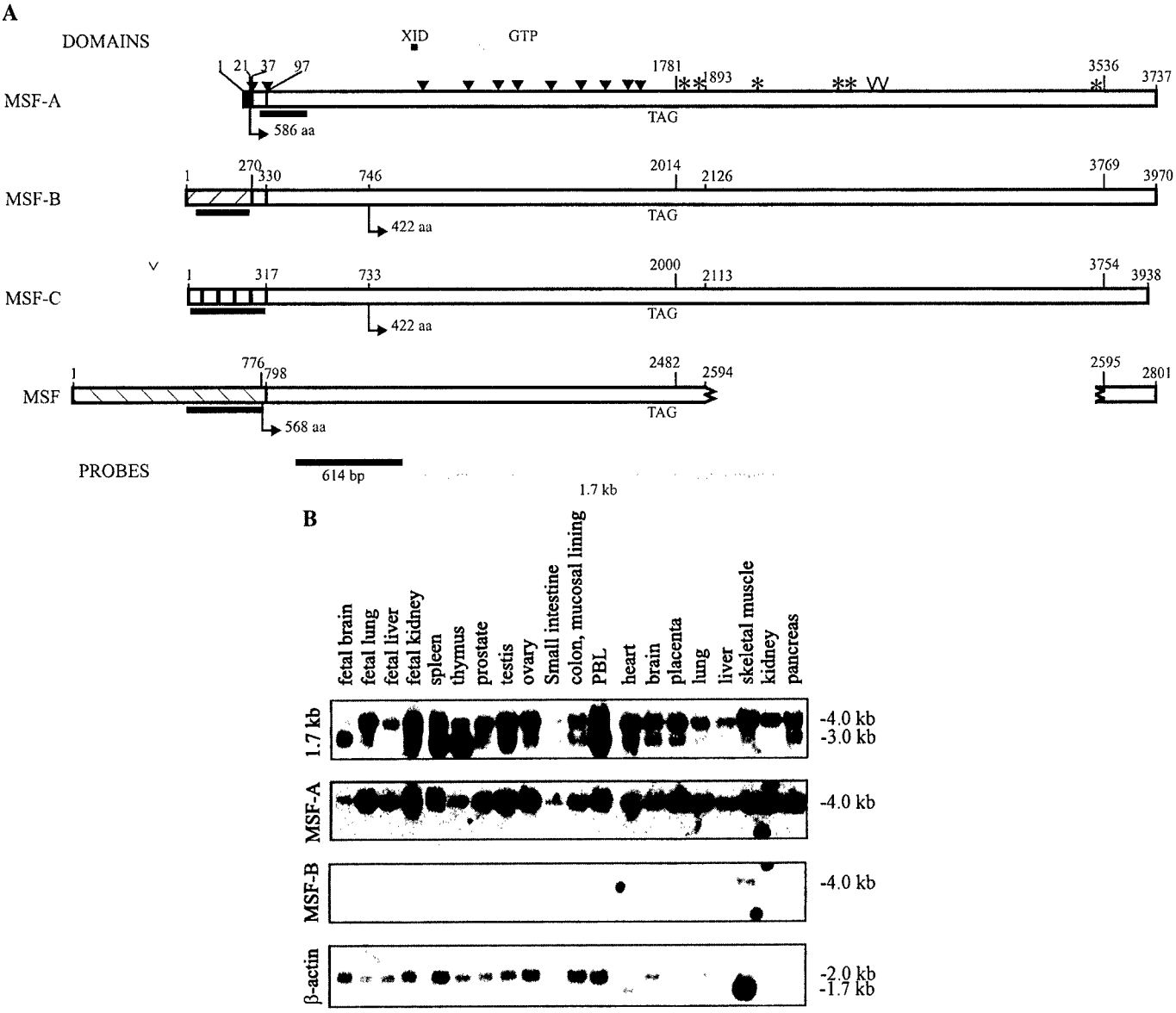


FIG. 1. Analysis of variant *MSF* transcripts. (**A**) Sequence diagram of variant *MSF* transcripts. Similar shading between variants indicates sequence identity. Black, front slash, vertical slash, and back slash boxes are unique 5′ regions. Black lines below each construct represent PCR amplified fragments for expression studies and *Not*I genomic clone mapping. ORFs are delimited at the start codon by an arrow and at the stop codon by the TAG. Numbers represent specific base positions in each variant sequence. Black triangles mark exon–exon boundaries; asterisks mark polymorphisms; "v" marks additional nucleotides in *MSF*-A and -B lacking in *MSF*-C (T/2637 *MSF*-A, 2870 *MSF*-B; G/2641 *MSF*-A, 2874 *MSF*-B). Black XID domain line indicates the xylose isomerase 1 domain; gray GTP domain line indicates position of the GTPase domain. Black probe line defines location of the 614-bp probe (Osaka *et al.*, 1999); gray probe line defines the location of the 1.7-kb *Eco*RI probe. (**B**) Northern blot autoradiograms of variant *MSF* expression. *MSF* probe locations are delimited in **A.** The 1.7-kb probe is the fragment common to all three variants defined as the gray probe line in **A.** PBL denotes peripheral blood leukocyte sample. β-actin signals in heart and skeletal muscle represent α and γ forms of actin. Probes were hybridized sequentially to the same set of blots.

(Figs. 1A and 2). All predicted ORFs began at a methionine located within a consensus Kozak sequence (Kozak, 1984) and terminated at a stop codon downstream of which was located an AATAAA polyadenylation site. *MSF*-A encoded a putative 586-amino-acid protein beginning at nucleotide 21 with a predicted molecular mass of 65.4 kDa. The first 20 residues of *MSF*-A were distinct from the first 7 amino acids of the published 568-amino-acid *MSF* protein sequence, after which the remaining sequences were identical. Both *MSF*-B and *MSF*-C had identical start methionines at nucleotides 746 and 733, respectively, from which a protein of 422 residues with a molecular mass of 47.5 kDa was predicted. An in-frame stop codon was located 189 amino acids upstream of the proposed start ATG in *MSF*-B and 208 amino acids upstream of the start ATG in *MSF*-C. The initiating methionine for *MSF*-B and *MSF*-C was found in-frame of both larger protein products from *MSF*-A and *MSF*. All predicted protein products contained the previously recognized conserved GTPase domain (Figs. 1A and 2). In addition, a xylose isomerase 1 domain was identified in all four variants

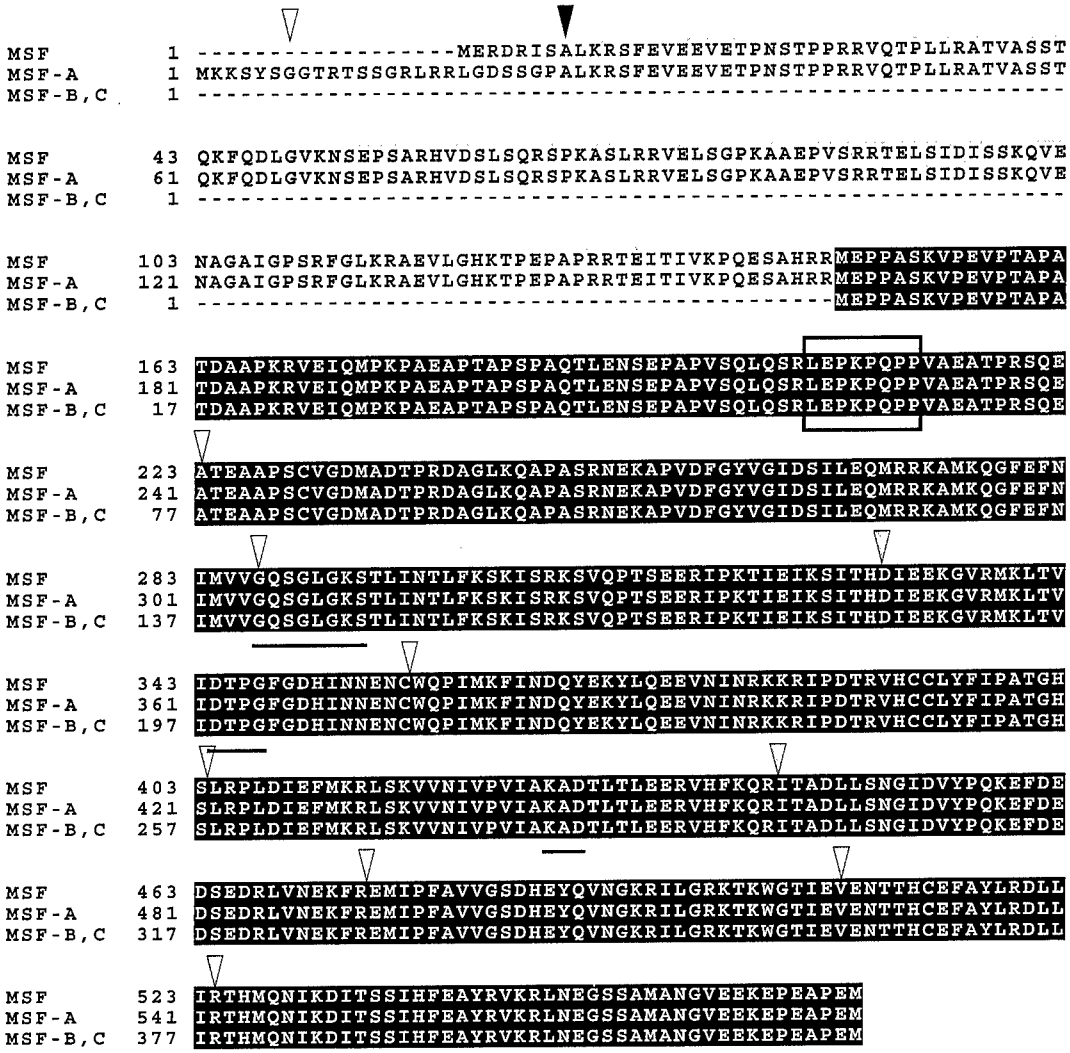


FIG. 2. Alignment of variant *MSF* predicted proteins. Gray background indicates common sequence between *MSF* and *MSF*-A. Black background indicates sequence common to all four variants. Gray triangles mark exon–exon boundaries. Black triangle marks exon–exon boundary that is breakpoint of *MSF* in *MLL–MSF* fusion protein (Osaka *et al.,* 1999). Numbering on left side indicates amino acids of each variant.

(Figs. 1A and 2). This highly conserved domain ([LI]-E-P-K-P-x(2)-P) is hypothesized to be important for catalysis and magnesium ligand binding in xylose isomerase, an enzyme that catalyzes the interconversion of specific sugars in microorganisms (Dauter *et al.,* 1989; Henrick *et al.,* 1989). A search of the databases using ScanProsite (http://www.expasy.ch/tools/scnpsit2.html) identified the FYN-binding protein (*FYB*), also known as the *SLP-76* associated protein (*SLAP-130*), as the only human protein to contain the xylose isomerase 1 domain. *FYB* is expressed specifically in T cells and myeloid cells and has been shown to be involved in the *FYN* and *SLP-76* signaling cascades (da Silva *et al.,* 1997; Musci *et al.,* 1997). Given that *MSF* was originally identified as part of a fusion protein in a t-AML patient, further work will be necessary to determine the functional significance of these highly conserved signaling domains. The prediction of three unique translation products from the *MSF* variant transcripts further supports the hypothesis that *MSF* plays multiple highly regulated roles in normal cellular metabolism.

Analysis of Variant MSF Transcript Expression

To investigate the adult and fetal expression pattern of *MSF*-A, -B, and -C, a 1.7-kb *Eco*RI fragment was hybridized against multitissue Northern blots (Fig. 1B). This fragment was completely contained within all three transcripts and partially overlapped the missing 1642-bp 3′ sequence in *MSF* (Fig. 1A). Variable expression of 4.0- and 3.0-kb transcripts in almost all adult and fetal tissues was detected (Fig. 1B). However, previous results by Osaka *et al.* (1999) using similar multitissue adult Northern blots showed ubiquitous expression of the 4.0-kb transcript and expression of the 3.0-kb transcript only in spleen, thymus, and peripheral blood leukocytes. In contrast, our data showed significant expression of the 3.0-kb transcript in fetal brain in addition to the hematopoetic tissues previously observed as well as near equal expression of the

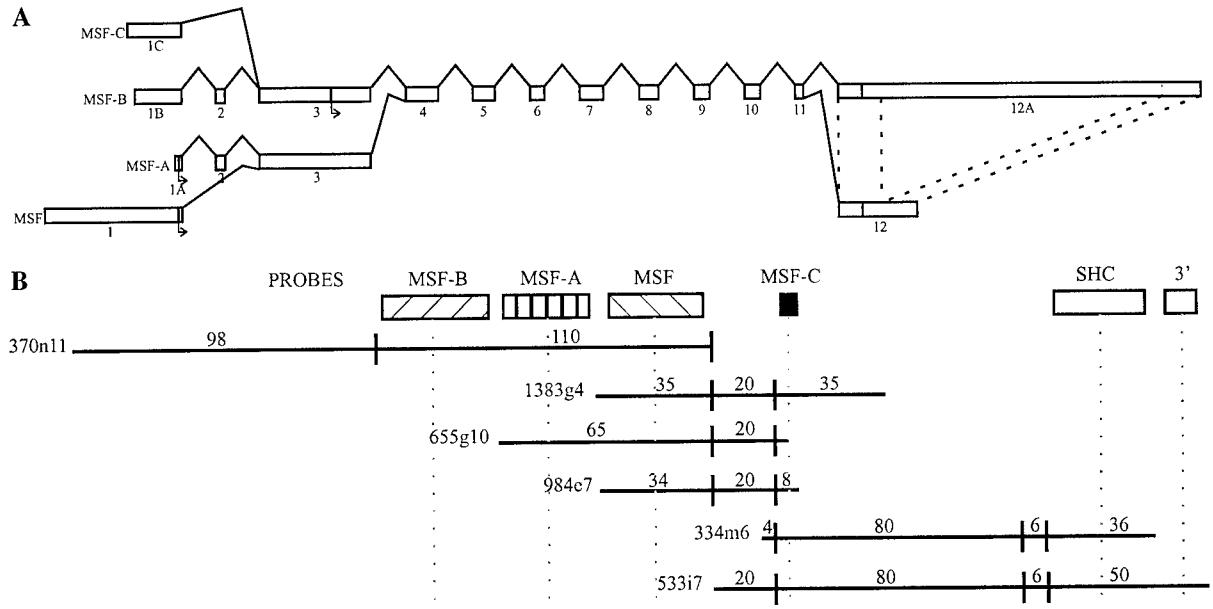


FIG. 3. Determination of variant *MSF* genomic structure. (**A**) Exon–intron boundaries. Boxes and connecting solid lines indicate exons and introns, respectively. The size of each exon is listed in Table 1. Exons are drawn to scale. Lengths of introns have not been determined except between exons 9 and 10 (257 bp). Dashed lines between exons 12 and 12A show noncontiguous sequences in exon 12A that are contiguous in exon 12 due to the absense of 1642 bp. (**B**) *Not*I genomic mapping. *Not*I sites are denoted by vertical lines. Sizes of *Not*I fragments are indicated in kilobases. Gene locations of *MSF* variant-specific 5′ probes are found in Fig. 1A. Boxes below probes indicate location of hybridization signal determined by overlapping *Not*I map. SHC probe defines the location of the original 339-bp septin-like solution hybrid capture cDNA fragment. The 3′ probe is a 594-bp PCR fragment amplified near the polyadenylation site. Vertical gray dotted lines aid in probe alignment with *Not*I fragments.

4.0- and 3.0-kb transcripts in other tissues including brain, testis, and small intestine. In addition, we failed to detect a previously observed transcript at 1.7 kb (Osaka *et al.*, 1999) in any tissues. It is unclear why our results differ from those in the initial *MSF* report. Given that we also used Clontech multitissue Northern blots, these inconsistencies may be related to variations in the stringency of hybridizations and washes. The probe for these previously published experiments was a 614-bp PCR fragment upstream and nonoverlapping with our 1.7-kb probe (Fig. 1A). However, our additional hybridizations with similar PCR products spanning this more 5′ region gave results consistent with those using the 1.7-kb *Eco*RI fragment (data not shown).

To characterize specific expression of each of the *MSF* transcripts, fragments representing the variable 5′ sequences were generated by RT-PCR (Fig. 1A). As the unique region of *MSF*-A spanned only 37 bp, no PCR fragment could be generated that was exclusively represented in *MSF*-A. A 245-bp amplified fragment that represented 7 bp unique to *MSF*-A and the remaining sequence common to *MSF*-A and *MSF*-B 5′ RACE products hybridized solely to the 4.0-kb transcript in all fetal and adult tissues (Fig. 1B). A 217-bp probe that was entirely unique to *MSF*-B identified a 4.0-kb transcript only in skeletal muscle (Fig. 1B). No signal was obtained with a 307-bp probe unique to *MSF*-C or a 315-bp probe unique to *MSF*. These results suggest that alternative splicing is involved in the expression of these transcripts, probably at the 5′ end. The 4.0- and 3.0-kb transcripts appear to be developmentally regulated given that the ratio of expression of the two transcripts changes from fetal brain to adult brain and from fetal kidney to adult kidney (Fig. 1B). As the isolated sequences of *MSF*-A and -B are close to the estimated length of the 4.0-kb observed Northern blot signal obtained with unique probes from these variants, it appears that these sequences are mostly complete. However, while *MSF*-B appears to be expressed in a tissue-specific manner, it is difficult to interpret these data completely as hybridization with a GAPDH probe (data not shown) and β-actin probe (Fig. 1B) suggested that the skeletal muscle mRNA sample was overloaded on this blot. The complete identity of the 3.0-kb transcript remains unclear as no sequences have yet been identified that hybridize solely to this transcript.

Determination of Genomic Structure and Identification of Polymorphisms

Evidence for the location of exon–intron boundaries was obtained by PCR amplification of 300- to 600-bp subfragments of *MSF* variants in total genomic DNA and epithelial cell line cDNA. Those intervals in which the expected PCR product was obtained in cDNA but was larger than expected or was unamplifiable in genomic DNA were sequenced in the corresponding BAC clone. Twelve total exons were identified ranging in size from 39 to 2091 bp (Fig. 3A). Sequence spanning

TABLE 1

Exon–Intron Splice Sequences and Polymorphisms

Exon	Acceptor splice site[a]	Donor splice site[a]	Size (bp)	Polymorphism	Restriction site
1	5′ UTR. . .	. . .CGG ATC TCA G**gt**acgcagac	797		
1a	5′ UTR. . .	. . .TCT TAC TCA G**gt**gggcttcg	39		
1b	5′ UTR. . .	. . .ctctgtccag**gt**gggtgcat	272		
1c	5′ UTR. . .	. . .ttcctgcaag**gt**aggccgaa	316		
2[b]	ctgtctttag**GA** GGC ACG CGG. . .	. . .AGT GGC CCA G**gt**aggtggct	56		
3	tatttttcag**CC** TTG AAA AGA. . .	. . .AGC CAG GAG G**gt**gagtcgca	643		
4	ccttcccca**gCC** ACT GAG GCG. . .	. . .ATG GTG GTC G**gt**gagtcctc	192		
5	ctctcccca**gGG** CAG AGC GGC. . .	. . .ATC ACG CAC G**gt**cagtggcc	128		
6	ctctgtgca**gAT** ATT GAG GAG. . .	. . .AAC GAG AAC TG**gt**gaggcccc	81		
7	ccccacgca**gC** TGG CAG CCC. . .	. . .ACC GGC CAC TC**gt**acgtccct	137		
8	tctgcttca**gC** CTC AGG CCC. . .	. . .AAA CAG CGG**gt**agggttcc	117		
9	cttatccca**gATC** ACC GCA. . .	. . .AAG TTC CGG**gt**gagtggat	95		
10	cgttcccca**gGAG** ATG ATC. . .	. . .ACC ATC GAA G**gt**actcgccg	96		
11	cctctcct**agTT** GAA AAC ACC. . .	. . .CTT CTC ATC AG**gt**gagagaca	51		
12	accccca**cagG** ACG CAC ATG. . .	. . .3′ UTR	454		
12a	accccca**cagG** ACG CAC ATG. . .	. . .3′ UTR	2091 (2072)[c]	I. acgcc(a/g)ccctgc[d]	G/*Aci*I,*Bso*FI
				II. aagtc(a/t)tttcc[d]	None
				III. ccgtg(c/t)gtgtg	None
				IV. ctgct(c/t)ggcca	T/*Bal*I
				V. cctgc(c/t)atccc	None
				VI. cttgg(c/t)ggcgg	C/*Bso*FI;T/*Dra*III

[a] Uppercase letters represent coding exonic sequences; lowercase letters represent intronic or noncoding exonic sequences. Conserved splice donor and splice acceptor dinucleotides are in boldface type.

[b] Exon 2 is translated in *MSF*-A and untranslated in *MSF*-B. Upper- and lowercase letters of this sequence represent *MSF*-A.

[c] Number in parentheses indicates length of last exon of the published sequence for *MSF*-C, which ends 19 bases earlier than *MSF*-A and *MSF*-B.

[d] Polymorphisms also found in exon 12 of *MSF*. All other polymorphisms are in the 1642-bp sequence unique to *MSF*-A, -B, and -C.

the exon–intron boundaries is listed in Table 1. All intron junction sequences contained the highly conserved splice donor GT and splice acceptor AG dinucleotides (Senapathy *et al.,* 1990). Reflecting the 5′ sequence variability, exon 1 was unique for each variant. Exons 1 and 1a spanned the proposed start sites for *MSF* and *MSF*-A, respectively, while exons 1b and 1c contained 5′ UTR sequences for *MSF*-B and *MSF*-C, respectively. Exon 2 was alternately spliced and found only in *MSF*-A as a coding exon and in *MSF*-B as a noncoding exon. Exons 3–11 were common to all variants. Exon 3 spanned the proposed start codon for *MSF*-B and *MSF*-C and therefore contained both untranslated and translated sequences for these variants. Exon 12 was unique to *MSF* as it lacked the 1642 nucleotides found in exon 12A of the remaining three variants. Six sequence polymorphisms all located in the 3′ untranslated region of exon 12A were identified (Table 1). Only polymorphisms I and II were also present in exon 12 of *MSF* as the remaining polymorphisms were located in the additional 1642 nucleotides of *MSF*-A, -B, and -C (Fig. 1A). Furthermore, polymorphisms I, IV, and VI were within nucleotide-specific restriction enzyme recognition sites.

From these genomic structure results, it was determined that the *MSF* breakpoint in the t-AML patient-derived *MLL–MSF* fusion protein (Osaka *et al.,* 1999) occurred at the splice junction between exon 1 and exon 3 (Fig. 2). Thus the fusion protein retained all but the first 7 residues of *MSF* in-frame including both the xylose isomerase 1 and the GTPase domains. The origin of the *MSF* variant-specific 1642 nucleotides in the 3′ UTR is unclear. PCR products generated from one primer positioned in the 5′ region common to all four sequences and from one primer positioned in the 1642-bp sequence were of equal size in human genomic DNA and cDNA in all samples studied (data not shown). Similarly, PCR products amplified from primers flanking the 1642 nucleotides were also of equal length in genomic DNA and cDNA, indicating that alternate splicing was not a contributing mechanism. A search of the EST database using the 1642-bp sequence plus an additional 50 bp of common flanking sequences identified numerous overlapping sequences spanning the interval. A search using the last 600 nucleotides of *MSF* did not identify any ESTs that spanned the breakpoint at nucleotide 2595 (Fig. 1A). However, several ESTs that contained partial sequences on either side of nucleotide 2595 were identified, suggesting that this region may be unstable.

To determine the genomic distance spanned by *MSF*, PCR-amplified subfragments of variants (Fig. 1A) were used as probes against a Southern blot of *Not*I digested genomic clones from the contig (Fig. 3B). Individual probes representing the variable 5′ sequences each gave unique *Not*I hybridization patterns, allowing for precise mapping. The *MSF*-B 5′ fragment mapped only to BAC 370n11 and furthest from the original 339-bp solution hybrid capture cDNA fragment on BAC 334m6. *MSF*-C mapped closest to the captured septin-like cDNA fragment while *MSF*-A and *MSF* unique 5′ PCR generated fragments localized between *MSF*-B and *MSF*-C. A 3′ PCR product amplified just proximal to the polyadenylation site did not hybridize to any of the previously isolated genomic clones in the contig and was hypothesized to map in one of the two gaps (Kalikin *et al.,* 1999). These primers were used to screen a total human BAC genomic library and yielded clone 533i7, which was determined by STS-PCR to overlap significantly with 334m6. *MSF* was therefore estimated to span 266 kb within the breast/ovarian candidate tumor suppressor gene region.

In summary, we describe three variant transcripts of the newly recognized *MSF* and note variant-specific tissue expression patterns. Analysis of the genomic structures provides evidence for 5′ alternative splicing in exon 1 that accounts for the majority of variation between transcripts. Subsequent to the submission of this paper, two additional AML patients were found with *MLL–MSF* fusion proteins (Taki *et al.,* 1999). Breakpoints in both patients occur at the identical sequence in *MSF* as initially described (Fig. 2; Osaka *et al.,* 1999). Thus, the identification of this gene as multiple fusion partners with *MLL* suggests that it may have activating potential while its localization to a discrete region of loss in breast and ovarian tumors suggests that it may be a candidate tumor suppressor gene. A review of the literature found that some disease genes such as *RET* are subject to both activating and inactivating mutations that lead to distinct clinical manifestations (Eng, 1999). Given the role GTPases have been shown to play in cellular proliferation (Bourne *et al.,* 1991) and the proposed role of the septin subfamily in cytokinesis (Kinoshita *et al.,* 1997), *MSF* merits further investigation as an attractive positional candidate for the 17q25 breast and ovarian tumor suppressor gene (Kalikin *et al.,* 1996, 1997).

ACKNOWLEDGMENTS

We thank C. K. Neeley and E. M. Bugeaud for technical assistance, S. Crider-Miller for technical advice on solution hybrid capture, E. R. Fearon for helpful discussion, and D. J. Law and K. R. Cho for critically reading the manuscript. This work was supported by NIH Grants F32CA71166-01 (L.M.K.), K08CA66613-01 (E.M.P.), and R01CA72877 (E.M.P.) and the Wendy Will Case Cancer Fund, Inc. (E.M.P.).

REFERENCES

Bourne, H. R., Sanders, D. A., and McCormick, F. (1991). The GTPase superfamily: Conserved structure and molecular mechanism. *Nature* **349:** 117–127.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Corral, J., Lavenir, I., Impey, H., Warren, A. J., Forster, A., Larson, T. A., Bell, S., McKenzie, A. N., King, G., and Rabbitts, T. H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: A method to create fusion oncogenes. *Cell* **85:** 853–861.

da Silva, A. J., Li, Z., de Vera, C., Canto, E., Findell, P., and Rudd, C. E. (1997). Cloning of a novel T-cell protein FYB that binds FYN and SH2-domain-containing leukocyte protein 76 and modulates interleukin 2 production. *Proc. Natl. Acad. Sci. USA* **94:** 7493–7498.

Dauter, Z., Dauter, M., Hemker, J., Witzel, H., and Wilson, K. S. (1989). Crystallisation and preliminary analysis of glucose isomerase from *Streptomyces albus*. *FEBS Lett.* **247:** 1–8.

Eng, C. (1999). RET proto-oncogene in the development of human cancer. *J. Clin. Oncol.* **17:** 380–393.

Fearon, E. R. (1997). Human cancer syndromes: Clues to the origin and nature of cancer. *Science* **278:** 1043–1050.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.

Futreal, P. A., Cochran, C., Rosenthal, J., Miki, Y., Swenson, J., Hobbs, M., Bennett, L. M., Haugen-Strano, A., Marks, J., Barrett, J. C., Tavtigian, S. V., Shattuck-Eldens, D., Kamb, A., Skolnick, M., and Wiseman, R. W. (1994). Isolation of a diverged homeobox gene, MOX1, from the BRCA1 region on 17q21 by solution hybrid capture. *Hum. Mol. Genet.* **3:** 1359–1364.

Henrick, K., Collyer, C. A., and Blow, D. M. (1989). Structures of D-xylose isomerase from *Arthrobacter* strain B3728 containing the inhibitors xylitol and D-sorbitol at 2.5 Å and 2.3 Å resolution, respectively. *J. Mol. Biol.* **208:** 129–157.

Jackson, R. J. (1993). Cytoplasmic regulation of mRNA function: The importance of the 3′ untranslated region. *Cell* **74:** 9–14.

Jean, J. C., Oakes, S. M., and Joyce-Brady, M. (1999). The Bax inhibitor-1 gene is differentially regulated in adult testis and developing lung by two alternative TATA-less promoters. *Genomics* **57:** 201–208.

Kalikin, L. M., Qu, X., Frank, T. S., Caduff, R. F., Svoboda, S. M., Law, D. J., and Petty, E. M. (1996). Detailed deletion analysis of sporadic breast tumors defines an interstitial region of allelic loss on 17q25. *Genes Chromosomes Cancer* **17:** 64–68.

Kalikin, L. M., Frank, T. S., Svoboda-Newman, S. M., Wetzel, J. C., Cooney, K. A., and Petty, E. M. (1997). A region of interstitial 17q25 allelic loss in ovarian tumors coincides with a defined region of loss in breast tumors. *Oncogene* **14:** 1991–1994.

Kalikin, L. M., George, R., Keller, M. P., Bort, S., Bowler, N. S., Law, D. J., Chance, P. F., and Petty, E. M. (1999). An integrated physical and gene mapping of human distal chromosome 17q24–proximal 17q25 encompassing multiple disease loci. *Genomics* **57:** 36–42.

Kinoshita, M., Kumar, S., Mizoguchi, A., Ide, C., Kinoshita, A., Haraguchi, T., Hiraoka, Y., and Noda, M. (1997). Nedd5, a mammalian septin, is a novel cytoskeletal component interacting with actin-based structures. *Genes Dev.* **11:** 1535–1547.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12:** 857–873.

Lavau, C., Szilvassy, S. J., Slany, R., and Cleary, M. L. (1997). Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J.* **16:** 4226–4237.

Musci, M. A., Hendricks-Taylor, L. R., Motto, D. G., Paskind, M., Kamens, J., Turck, C. W., and Koretzky, G. A. (1997). Molecular cloning of SLAP-130, an SLP-76-associated substrate of the T cell antigen receptor-stimulated protein tyrosine kinases. *J. Biol. Chem.* **272:** 11674–11677.

Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1999). Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* **6:** 63–70.

Osaka, M., Rowley, J. D., and Zeleznik-Le, N. J. (1999). MSF (Mll septin-like fusion), a fusion partner gene of MLL, in a therapy-related acute myeloid leukemia with a t(11;17)(q23;q25). *Proc. Natl. Acad. Sci. USA* **96:** 6428–6433.

Phelan, C. M., Borg, A., Cuny, M., Crichton, D. N., Baldersson, T., Andersen, T. I., Caligo, M. A., Lidereau, R., Lindblom, A., Seitz, S., Kelsell, D., Hamann, U., Rio, P., Thorlacius, S., Papp, J., Olah, E., Ponder, B., Bignon, Y. J., Scherneck, S., Barkardottir, R., Borresen-Dale, A. L., Eyfjord, J., Theillet, C., Thompson, A. M., Devilee, P., and Larsson, C. (1998). Consortium study on 1280 breast carcinomas: Allelic loss on chromosome 17 targets subregions associated with family history and clinical parameters. *Cancer Res.* **58:** 1004–1012.

Rowley, J. D. (1998). The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* **32:** 495–519.

Saito, H., Inazawa, S., Kasmuni, F., Koi, S., Sagae, S., Kudo, R., Saito, J., Noda, K., and Nakamura, Y. (1993). Detailed deletion mapping of chromosome 17q in ovarian and breast cancers: 2-cM region on 17q21.3 often and commonly deleted in tumors. *Cancer Res.* **53:** 3382–3385.

Senapathy, P., Shapiro, M. B., and Harris, M. L. (1990). Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to Genome Project. *Methods Enzymol.* **183:** 252–278.

Swaroop, A., and Xu, J. (1993). cDNA libraries from human tissues and cell lines. *Cytogenet. Cell. Genet.* **64:** 292–294.

Taki, T., Ohnishi, H., Shinohara, K., Sako, M., Bessho, F., Yanagisawa, M., and Hayashi, Y. (1999). AF17q25, a putative septin family gene, fuses the MLL gene in acute myeloid leukemia with t(11;17)(q23;q25). *Cancer Res.* **59:** 4261–4265.

Theile, M., Hartman, S., Scherthan, H., Arnold, W., Deppert, W., Frege, R., Glaab, R. F., Haensch, W., and Scherneck, S. (1995). Suppression of tumorigenicity of breast cancer cells by transfer of chromosome 17 does not require transferred BRCA1 and p53 genes. *Oncogene* **10:** 439–447.